[The effect of zonal factors in estimating crash risks by transportation modes: Motor vehicle, bicycle and pedestrian](https://www.researchgate.net/publication/309308779_The_effect_of_zonal_factors_in_estimating_crash_risks_by_transportation_modes_Motor_vehicle_bicycle_and_pedestrian?enrichId=rgreq-3b4ab993223fa5ce673bd2aa7be00120-XXX&amp;enrichSource=Y292ZXJQYWdlOzMwOTMwODc3OTtBUzo1NDc3OTY1NDM4MzYxNjBAMTUwNzYxNjI4MTY0NQ%3D%3D&amp;el=1_x_3&amp;_esc=publicationCoverPdf)

**The Effect of Zonal Factors in Estimating Crash Risks by Transportation**

**Modes: Motor Vehicle, Bicycle and Pedestrian**

Jie WANG1, Helai HUANG1\*, Qiang ZENG2

1 School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan, China

2 School of Civil Engineering and Transportation, South China University of Technology, Guangzhou,

Guangdong, China

**Abstract:** *Objectives:* This paper aimed to (i) differentiate the effects of contributory factors on

crash risks related to different transportation modes, i.e., motor vehicle, bicycle and pedestrian; (ii)

explore the potential contribution of zone-level factors which are traditionally excluded or

omitted, so as to track the source of heterogeneous effects of certain risk factors in

crash-frequency models by different modes. *Methods:* Two analytical methods, i.e. negative

binomial models (NB) and random parameters negative binomial models (RPNB), were

employed to relate crash frequencies of different transportation modes to a variety of risk factors

at intersections. *F*ive years of crash data, traffic volume, geometric design as well as macroscopic

variables at traffic analysis zone (TAZ) level for 279 intersections were used for analysis as a case

study. *Results:* Among the findings are: (1) the sets of significant variables in crash-frequency

analysis differed for different transportation modes; (2) omission of macroscopic variables would

result in biased parameters estimation and incorrect inferences; (3) the zonal factors (macroscopic

factors) considered played a more important role in elevating the model performance for

non-motorized than motor-vehicle crashes; (4) a relatively smaller buffer width to extract

macroscopic factors surrounding the intersection yielded better estimations.

**Keywords:** transportation modes; macroscopic variables; unobserved heterogeneity; buffer

width; intersection safety

# Introduction

Many communities have increased their interest in the implementation of multimodal

transportation and advocated for the shift from motor vehicles to non-motorized modes of

transportation, i.e., walking and cycling. In spite of the health and environmental benefits, an

increasing number of crashes involving pedestrians and bicyclists has become a major concern in

improving traffic safety. For example in 2013, the United States had 4735 pedestrian and 743

bicyclist deaths, accounting for 18% of all U.S. highway fatalities (NHTSA, 2013). The Federal

Highway Administration’s office of safety has established pedestrian and bicyclist safety as one

of its top priorities. Thus, it is essential for traffic safety engineers to provide appropriate

countermeasures or policies to achieve friendly and safe multimodal transportation.

A comprehensive understanding of contributing factors associated with crash occurrences by

different modes is a prerequisite for developing safety improvement programs to effectively

reduce traffic crashes. For a given road entity (e.g. road segments or intersections), the potential

factors associated with multimodal crashes could be summarized as in Figure 1, according with

Miranda-Moreno et al. (2011), Mitra and Washington (2012), Ukkusuri et al. (2012) and Strauss et

al. (2003; 2014). The factors influencing road-entity-level crash frequency by modes include

macroscopic factors related to built environment of the road entities - such as population and

economic characteristics, land use characteristics and travel behaviors - as well as road features

and traffic characteristics of the road entities. In addition, crash occurrence is also associated with

individual characteristics such as gender, age, education, alcohol consumption, and other driver

and pedestrian behaviors (Ryb et al., 2007). Although discrete individual-level factors are not

available to be integrated into the crash-frequency model, individual characteristics are always

influenced by macroscopic factors (Christoffel and Gallagher, 1999). Therefore, macroscopic

factors could serve as a surrogate for individual behaviors.

The choice of appropriate analytical method and the selection of representative explanatory

variables are two important considerations for obtaining accurate model predictions. Over the

past three decades, considerable research efforts have been devoted to developing and applying

sophisticated methodological approaches associated with the analysis of crash frequency.

Detailed descriptions and assessments of crash-frequency models can be found in the review

papers by Lord and Mannering (2010) and Mannering and Bhat (2014). However, relatively few

studies have focused on the identification and inclusion of traditionally excluded or omitted

variables in crash-frequency analysis. In particular, variables related to macroscopic factors

previously described (in Figure 1) are normally unavailable in crash databases and as a result

have rarely been examined in great detail. Mitra and Washington (2012) is one of a few studies

exploring the omitted variables in crash-frequency modeling. The authors developed two

different models of estimating intersection crash frequency, one with traffic volume as the only

independent variable, and the other with several spatial factors in addition to commonly

included geometric design and traffic factors. Through contrastive analysis of the two models,

results indicated that some spatial factors, such as local influences of weather, sun glare,

proximity to drinking establishment, proximity to school and demographic attributes near

intersections, have significant explanatory power and their exclusion leads to biased estimates.

Statistical methods such as spatial and temporal correlation, multilevel, random effect,

random parameter, and latent class approaches have been developed to address this issue of

unobserved heterogeneity (Anastasopoulos and Mannering, 2009; Dong et al., 2016; Mannering et

al., 2016; Quddus, 2008; Wang and Huang, 2016; Xu and Huang, 2015; Xu et al. 2016), as these

omitted explanatory variables can be regarded as part of the unobserved heterogeneity.

Unobserved heterogeneity impacts traffic safety analysis in two ways: the first problem is that the

selected explanatory variables cannot fully account for the cross-section or longitudinal-section

variations in crash counts due to unobserved road geometrics, environmental factors, driver

behavior and other confounding factors, which lead to impaired predictive performance of the

model (called heterogeneity in model prediction); the second problem is that these unobserved

factors are always correlated with observed factors and thus biased parameters will be estimated

and incorrect inferences could be drawn (called heterogeneity in the coefficient estimator). While

these approaches will mitigate the adverse impacts of omitting significant explanatory variables,

the resulting model estimates still fail to track the original source of heterogeneity and quantify

the safety effect of omitted variables (such as macroscopic factors shown in Figure 1). Omission of

important explanatory variables still remains a problem even with advanced statistical

approaches to capture unobserved heterogeneity (Mannering et al., 2016).

The study by Mitra and Washington (2012) attempted to investigate the safety effect of some

important omitted variables on total crash frequency and their contribution on model estimation.

As Venkataraman et al. (2013) stated, frequency models of crash outcome type can provide

substantial insights into the effect of explanatory variables and assist in examining the

heterogeneity effects in roadway geometric features. This paper aims to extend previous research

(Mitra and Washington, 2012) and investigate how macroscopic factors affect the crash-frequency

analysis for different transportation modes. This is because there may be some inconsistent

impacts of some macroscopic variables on motor vehicle and non-motorized (including bicycle

and pedestrian) crashes. For example, Lee et al. (2015) utilized a multivariate model for

investigating motor vehicle and non-motorized crashes at the macroscopic level. Results for the

parameter estimation suggested that some zonal variables related to demographics and road

characteristics have different directional effects on motor vehicle and non-motorized crashes.

Meanwhile, the most appropriate width of buffer to extract macroscopic factors may be

inconsistent between modeling motor vehicle and non-motorized crashes. Therefore, it is

advisable to model the crash frequency by separate transportation modes to examine the effects

of macroscopic factors.

In summary the objective of this paper is twofold: (1) to examine the effects of a host of

contributing factors including both macroscopic and microscopic factors on crash occurrence

with respect to different transportation modes; (2) to shed further light on the contribution of the

macroscopic factors which are traditionally excluded or omitted variables, to tracking the source

of heterogeneity effects in coefficient estimators of regularly used variables and improving the

model performance in crash-frequency analysis related to different modes.

## 2 Data preparation

In this study, data collected for 279 intersections located in Hillsborough County, Florida,

USA were used to develop the intersection crash-frequency models for different transportation

modes. The data for the analysis was mainly divided into four types: traffic crash data, traffic

characteristics, road characteristics related to geometric design, traffic control/regulatory of the

intersection, macroscopic factors including trip production/attraction, demographic and

socio-economic characteristics surrounding the intersection. The derivation and processing of

these data sources are described next.

*2.1 Crash data*

Crash data for the intersections in a five-year period (2005–2009) were obtained from the

Florida Department of Transportation (FDOT) Crash Analysis Reporting (CAR) system. Crashes

were categorized as intersection related crashes if they occurred within the curb-line limits of

the intersection or if they occurred within the influence area of the intersection, which is 250 feet

away from the stop line. Intersection-level crash data was disaggregated into motor vehicle,

bicycle, and pedestrian crashes. A motor vehicle crash was defined as a collision between two or

more motor vehicles or between a motor vehicle and an object. A bicycle crash referred to a

collision between a motor vehicle and a bicycle. Likewise, a pedestrian crash denoted a collision

between a motor vehicle and a pedestrian.

*2.2 Traffic characteristics*

Previous researches suggest that traffic characteristics such as motor vehicle, pedestrian and

bicycle volume are the most important factors influencing crash occurrences. Motor vehicle

volume represented by average annual daily traffic (AADT) can be collected from the FDOT

Roadway Characteristics Inventory. Two motor vehicle volume variables including AADT from

major road and AADT from minor road of 5-year (2005–2009) average were also obtained. Actual

pedestrian and bicycle volume are not regularly available. The collected macroscopic data such as

population were used to serve as a surrogate for pedestrian and bicycle volume as suggested by

Jacobsen (2003) and Miranda-Moreno et al.(2011).

*2.3 Road features*

Road features related to geometric design and regulatory/control attributes of the road entity

were collected from the FDOT Roadway Characteristics Inventory. The road factors considered in

the study are number of legs, presence of traffic signal, speed limit on major approach, and speed

limit on minor approach.

*2.4 Macroscopic factors*

Considerable previous studies on zonal-level crash-frequency models suggests that various

macroscopic factors such as trip production/attraction, demographic and socio-economic

characteristics affect area-wide traffic crashes (Quddus., 2008; Huang et al., 2010, 2016; Abdel-Aty

et al., 2011; Xu et al., 2014; Dong et al., 2015, 2016). It is hypothesized that the number of crashes

occurring at an intersection is also associated with these macroscopic factors surrounding the

intersection.

Trip production/attraction factors such as total trip productions/attraction, home-based work

productions/attraction, college productions/attraction at the TAZ level were collected from the

Intermodal Systems Development Unit of District 7 of the FDOT. Demographic and

socio-economic characteristics were examined including the geographical area of each TAZ,

population, income and commuting, which were downloaded from the United States Census

report.

ArcGIS 10.0 was used to generate a buffer around each selected intersection and conduct a

spatial analysis to extract macroscopic data from the TAZ layers. The process is described in

detail as follows. First, a spatial overlay of TAZ layers on a specified width buffer (including

0.25 mile, 0.5 mile and 1 mile buffer) was generated around each intersection. Then, spatial

analysis operators such as “intersect” and “join” available in the GIS environment were used to

intersect layers, join tables and extract selected trip production/attraction, demographic and

socio-economic characteristics within the generated buffer. Macroscopic factors were distributed

in proportion to the area of TAZ within the generated buffer. The ArcGIS procedure adopted to

extract and estimate macroscopic factors in this study was similar to the one discussed in detail

by Pulugurtha and Sambhara (2011) to develop pedestrian crash estimated models.

Table 1 provides descriptive statistics of crash data, traffic variables, road variables and

macroscopic variables located in 0.5 mile buffer. The values of macroscopic variables located at

0.25 and 1 mile buffer are not listed for compactness of the table.

## 3 Methodology

In previous crash-frequency analyses, Poisson and Negative binomial model (NB), along

with their variants (such as Poisson-lognormal model), are commonly used and proven to be

successful as they effectively model the rare, random, sporadic, and non-negative crash data. As

crash data exhibit over-dispersion (i.e., variance greater than mean), NB is superior to the Poisson

model. Compared with the basic Poisson model, NB includes a gamma-distributed error term in

Poisson mean to account for the over-dispersion due to omission of relevant variables or

measurement error in crash data. The formulation for NB can be presented as follows:

11

Where *i*s the crash frequency by modes (i.e., motor vehicle, bicycle and pedestrian) at

intersection, and is the expectation ofis a vector of explanatory variables.

intercept,is a vector of estimable parameters. is the error term that is assumed to be

independentand has a two-parameter gamma distribution.

The NB model presented in Eq.(2) could control for unobserved heterogeneity by omitted

variables. However, this model assumes that the unobserved variables are uncorrelated with the

observed exploratory variables. If this correlation exists, unobserved factors can introduce

variation in the effect of observed variables on crash likelihood. Random parameters approaches

are able to address this issue by allowing non-constant estimable parameters to vary across

observations (Mannering et al., 2016). In random parameters negative binomial model (RPNB), estimable parameters in Eq.(2) can be written as:

3

Whereis the mean of the random parameter is a randomly distributed term (e.g.,

a normally distributed term with mean 0 and variance ) that capture heterogeneity across

**

observations. The analyst can test for random parameters with Eq.(3), across all observations i for

each included explanatory variable. If the variance of the chosen distribution is not significantly

different from zero, it suggests that a conventional fixed parameter is statistically appropriate.

Thus the model always combines fixed and random parameters across the included explanatory

variables.

As previously stated, heterogeneity effects of certain risk factors mainly derive from the

combined effects of unobserved variables that have been omitted from the model. Although the

random parameters approaches could mitigate the adverse impacts of omitting variables, the

original source of unobserved heterogeneity (or what are the major factors that lead to

unobserved heterogeneity) still fails to be well understood. Thus one of the aims of this study is

to test the potential role of macroscopic variables (referring to as the influential omitted variables)

in tracking the source of heterogeneity effects related to commonly used traffic and road

variables.

To this end, two different model specifications are estimated and compared, both with

random parameters approaches. In the base model only traffic and road variables are included,

while the second model traffic and road variables as well as macroscopic variables at TAZ level

surrounding the intersections are included. If the test results for parameter estimation on a

variable appear as random in base model (variance of the chosen distribution is significantly

different from zero), while this variable has the fixed effect in the second model. Then we could

infer that the heterogeneity effect of this variable is mainly caused by these macroscopic variables.

For another case, the safety effect of this variable is still random in second model; the source of

heterogeneity effect of this variable still cannot be clearly distinguished, maybe due to other

important omitted variables.

Apart from the potential role in accounting for the heterogeneity effects in parameters

estimation, integrating the macroscopic variables could also decrease the variance of the random

error (i.e. overdispersion) and thus improve the model performance in crash-frequency

prediction. The proportion of reduction in variance (PRV), also called explained variance,

proposed by Raudenbush and Bryk (2002) can be used to assess the overall explanatory power of

macroscopic factors for modeling the crashes by different modes. In this case, the PRV of

macroscopic variables is defined as:

Where is the variance of the error term in the base model without macroscopic variables.

is the variance of error term in the full model with macroscopic variables. The value of PRV is

bounded by 0 and 1, and a higher value indicates a stronger explanatory power of macroscopic

factors on the crash occurrence.

Furthermore, two goodness-of-fit statistics are used for model comparisons: Akaike

Information Criterion (AIC) and log-likelihood ratio (LR). The AIC is calculated as follows:

44

Where LL is the log-likelihoods at convergence for the estimated model, andis the

number of parameters in the statistical model. The model with the lower AIC is considered to

have the better goodness of fit.

The LR value is the chi-squared value in the log-likelihood ratio test for the null hypothesis

test that reveals whether or not the equivalence of two models should be rejected. The likelihood

ratio statistic is,

which isdistributed with J degrees of freedom, where J = KA − KN (KA and KN are the

number of coefficients for the alternative model and the null model, respectively ), LLN and LLA are

the log-likelihoods at convergence for the null model and the alternative model, respectively. The

null hypothesis for Eq. (6) is that the alternative model does not have a significantly lower

log-likelihood than the null models, indicating a lack of significant difference between the null

model and the alternative model.

## Results and discussion

Three types of crash-frequency models for motor vehicles, bicycles and pedestrians were

developed. Each type of model involved eight separate models based on four model

specifications (one with only traffic volume and road features, the other three with macroscopic

factors overlaid on 0.25, 0.5, 1 mile buffer respectively in addition to commonly included traffic

volume and road features ) and two analytical methods (NB and RPNB).

LIMDEP econometric software was used to develop the statistical models described above.

To enable focus on the most significant variables, variables that were not found to significantly

different from zero at the 0.1 level of significant using a *t*-test were removed. Meanwhile, the

likelihood ratio test was used to guarantee that each added variable significantly improved the

overall model performance. In the RPNB, if the variance of a random parameter was not

statistically different from zero, the random parameter was simplified to be fixed across

intersections. Thus, the results in NB were in accordance with that in RPBN when no estimate

parameters of explanatory variable were statistically random.

This analysis below will emphasize testing effects of macroscopic factors on model

performance in crash-frequency analysis for three transportation modes, and then comparison

results of the parameter estimates and marginal effects between the base model and the full

model with macroscopic variables will be presented and interpreted.

*4.1 Effects of macroscopic factors on model performance*

Tables 2-4 show goodness-of-fit measures for motor vehicle, bicycle and pedestrian

crash-frequency models, respectively. As shown in Table 2, only five of eight motor vehicle

crash-frequency models, including two base models (NB model and RANB model) and three

fully specified NB models, were presented since there was no significant random parameter as

measured by the t-statistics in all three models with macroscopic variables. Although there was

no substantial difference in goodness-of-fit as reflected by likelihood ratio test between the base

NB and RPNB model, the present of significant random parameters ( e.g. the variable of ‘presence

of traffic signal’ in this case study) demonstrated the existent of heterogeneity of risk factors in

base model without considering macroscopic factors. More interestingly, no significant random

parameters were found in all three full models with macroscopic variables. This implied that the

heterogeneous effects of risk factors on motor vehicle crash frequency could be mostly captured

by these macroscopic variables, at least for the Hillsborough dataset examined here. Frequency

analysis models for bicycle crashes presented a similar result to motor vehicles crashes, as shown

in Table 3. However, this was not the case for the pedestrian crash-frequency models (Table 4).

Significant random parameters, such as ‘presence of traffic signal’, existed both in pedestrian

crash models with and without macroscopic variables, suggesting that the heterogeneity effect in

parameters estimation cannot be completely picked up by these macroscopic variables.

Apart from the potential effect in tracking the heterogeneity, results also revealed that

incorporating the macroscopic variables in crash-frequency analysis leaded to an increasing

model complexity but a considerable improvement in overall fit as measured by log likelihood at

convergence. As shown in Table2, the likelihood ratio test comparing the full NB models and the

base NB models indicated that we were more than 99.99 % confident that the full models with

macroscopic variables (except the full model with 1.0 mile buffer-width macroscopic variables)

were statistically superior. This comparison suggested that the macroscopic variables explained a

portion of variability in crash occurrences and should not be omitted in motor vehicle

crash-frequency model. In regard to bicycle and pedestrian models, omission of macroscopic

variables will also lead to a significant decrease in goodness-of-fit, as shown in Tables 3-4.

Comparing model outputs developed based on 0.25 mile, 0.5 mile and 1 mile buffer width

data, these models with macroscopic variables of 0.25 mile buffer width had the lowest AIC,

conversely, the models with macroscopic variables of 1 mile buffer width had the highest AIC in

all three types of crash-frequency models by modes. Thus, a relatively smaller buffer width in

extracting macroscopic factors surrounding the intersection would yield a better estimate.

To further assess and compare the overall explanatory power of macroscopic variables, the

values of PRVs were calculated. As shown in Table 2, the motor vehicle crash-frequency model

with macroscopic variables of 0.25 mile buffer width had the highest PRV of 7.98%. This meant

that 7.98% of unexplained variation resulted from those omitted macroscopic variables, which

also suggested the usefulness of the motor vehicle crash-frequency analysis by integrating

macroscopic factors. Accordingly, the highest values of PRV were 33.02% and 26.37% in bicycle

and pedestrian crash-frequency models respectively, as shown in Tables 3-4.

By comparing PRVs in models by different transportation modes, the PRVs in bicycle and

pedestrian crash-frequency model were much higher than that in motor vehicle models. In other

words, integrating macroscopic factors in non-motorized crash-frequency model was more vital

than that in developing motor vehicle model. This result was in line with the expectation. One

possible reason for this distinct effect is that pedestrian/bicycle volume (or pedestrian/bicycle

activity) which is commonly identified as the main determinants of pedestrian/ bicycle crash

frequency has been omitted in the base pedestrian/bicycle crash-frequency models. Integrating

macroscopic factors for pedestrian/bicycle crash-frequency analysis made up the absence of

pedestrian/bicycle volume in predicting pedestrian/bicycle crash frequencies to some extent as

demonstrated by previous study (Jacobsen, 2003; Miranda-Moreno et al., 2011) that macroscopic

data can serve as a surrogate for pedestrian and bicycle volume. Another reason, maybe even

more importantly, originates from the differences in the travel distance between non-motorized

and motor vehicle modes. As walking and bicycle are short-distance transportation modes, crash

victims of pedestrians and bicyclists generally reside near the crash intersection, and thus, the

macroscopic factors extracted surrounding the intersection can probably better reflect pedestrian

and/or bicyclist behaviors than that for motor drivers.

*4.2 Parameter estimates and marginal effects*

Three types of crash-frequency models for motor vehicles, bicycles and pedestrians were

estimated and each type involved eight separate models based on four model specifications and

two analytical methods, yielding a total of 24 models. Three full models which have a better

goodness-of-fit, including motor vehicle NB model with 0.25 mile buffer width macroscopic

variables, bicycle NB model with 0.25 mile buffer width macroscopic variables and pedestrian NB

model with 0.25 mile buffer width macroscopic variables, were selected as recommended models

for reasons outlined. Meanwhile, the parameter estimates of three base NB models for motor

vehicles, bicycles and pedestrians were presented for comparison. Tables 5-7 show the parameter

estimates and their t-statistics for motor vehicle, bicycles and pedestrians crash-frequency models,

respectively. Since the model has nonlinear coefficients, direct parameters will not show a unit

effect on the number of crashes. Table 8 thus summarizes the results for marginal effects, which

could be interpreted as the average impact of a unit change in an explanatory variable on crash

frequency.

Comparing marginal effects of microscopic variables (traffic volume and road variables)

between the full NB models and the base NB models revealed some important differences. The

major differences were in marginal estimates of the variable ‘presence of traffic signal.’ The

results in the base model using NB approach showed that ‘presence of traffic signal’ had positive

association with the crash frequency for all three transportation modes; while this variable

became no statistically significant for motor vehicle and bicycle crashes in the full NB models.

Meanwhile, the present of macroscopic variables also modified the marginal effects of

microscopic variables (see Table 8). For example, the marginal effects of ln(AADT-minor) were

6.38 and 0.14 respectively for motor vehicle and bicycle crashes in the base models, while these

values were 7.67 for motor vehicle crashes and 0.17 for bicycle crashes in models with

macroscopic variables. This difference clearly showed that in the absence of important

macroscopic variables, the marginal effects of ln(AADT-minor) for motor vehicle and bicycle

crashes were biased downwards by 16.8% and 17.6% respectively. These results agreed with the

safety research by Mitra and Washington (2012) that the exclusion of important variables may

cause bias in coefficient estimates and incorrect inferences.

According to the results of parameter estimates (Tables 5-7) and their marginal effects (Table

8), significant variable sets for crashes were not consistent for different transportation modes.

‘AADT on major approach’ and ‘density of total population’ were two contributing factors that

had statistically significant effects on the three response variables (i.e., motor vehicle, bicycle and

pedestrian crash frequency). Four variables including ’AADT on minor approach,’ ‘number of

legs’, ‘proportion of college productions and attractions’ and ‘proportion of workers commuting

by public transportation’ were significant for two response variables. Six variables were solely

associated with one response variable: ‘presence of traffic signal’, ‘speed limit on major

approach,’ ‘speed limit on minor approach,’ ‘proportion of home-based productions and

attractions,’ ‘proportion of population between age 16 and 64’ and ‘proportion of workers

commuting by walking’. The detailed interpretations for these significant risk factors are offered

in the following.

*4.2.1 Traffic volume*

Similar to numerous prior studies, traffic volumes are significant variables for intersection

crashes and are positively correlated with crash occurrence (Lee and Abdel-Aty, 2005; Mitra and

Washington, 2012; Xie et al., 2013). The marginal effects of ln(AADT-major) were 47.65 , 0.59 and

1.19 respectively for motor vehicle, bicycle and pedestrian crashes in the full models, indicating

that an average of thousand increase in major approach AADT will lead to a 2.31, 0.03 and 0.06

increase in motor vehicle ,bicycle and pedestrian crash frequency respectively. Similarly, an

average of thousand increase in minor approach AADT was associated with a 0.69 and 0.02

increase in motor vehicle and bicycle crashes.

*4.2.2 Number of legs*

The marginal effects of ‘number of legs’ on motor vehicle and bicycle crashes were 24.28 and

0.43. This result suggested that the four-legged intersection was associated with 24.28 more motor

vehicle crashes and 0.43 more bicycle crashes compared to the intersection with three legs. This

result was generally expected and agreed with the preliminary finding that a larger number of

legs may increase the likelihood of crash occurrence due to more potential conflicts (Zeng and

Huang, 2014). However, this variable did not found to have significant effects on pedestrian crash

frequency. This may be due to the higher design standards of facilities (such as marked

crosswalks) in the large-leg intersection that leads to mixed effects of ‘number of legs’ on

pedestrian safety.

*4.2.3 Traffic signal*

The effects of traffic signal on safety are very interesting. The results of parameter estimates

in the base model using NB approach showed that ‘presence of traffic signal’ had positive

association with all three target variables. This is not consistent with the empirical hypothesis that

the installation of traffic lights could improve intersection safety. In the base model using RPNB

approach, ‘presence of traffic signal’ had a positive effect on three target variables but with a

varying magnitude across intersections (the results for the RPNB model were not presented since

there was no significant improvement in goodness-of-fit compared to the NB model). More

interestingly, this variable became no statistically significant for both motor vehicle and bicycle

crashes in the models integrating macroscopic variables. The possible reason for this difference is

that the macroscopic variables could account for a positive and heterogeneity effect of the

‘presence of traffic signal’. This implies that the installation of traffic lights itself will not increase

the crash risk but traffic lights are always installed at relatively hazardous sites; however, more

work is needed to verify this conclusion.

*4.2.4 Speed limit*

The variables related to speed limit had inconsistent effects on crash occurrence by different

modes. ‘Speed limit on minor approach’ was positively correlated with the frequency of motor

vehicle crashes. An increase of 10 mph in speed limit on minor approach will increase motor

vehicle crashes by 18.7. This is generally expected since at high speeds the time to react to

changes in the environment is shorter, leading to higher crash frequency. However, ‘speed limit

on major approach’ was negatively associated with pedestrian crashes. The frequency of

pedestrian crashes will decrease by 0.6 with per 10 mph increase in speed limit on major

approach. The probable reason for this result is that a higher speed limit is always related with

higher design standards of facilities such as pedestrian overcrossing and underpass. The effect of

speed limit on bicycle crashes was not significant.

*4.2.5 Trip characteristics*

Results showed that the proportion of home-based productions and attractions near an

intersection was negatively associated with motor vehicle crashes. The frequency of motor

vehicle crashes will decrease by 1.32 with a percentage increase in proportion of home-based

trips. This is reasonable since drivers of a home-based trip are more familiar with the traffic

environment and have more cautious driving behaviors (Abdel-Aty et al., 2011). ‘Proportion of

college productions and attractions’ was negatively associated with motor vehicle and bicycle

crashes while was not statistically significant for pedestrian crashes. A percentage increase in

proportion of college trips will result in a 1.52 and 0.03 decrease in motor vehicle and bicycle

crashes. The decreased motor vehicle and bicycle crashes may be due to better traffic control

measures in these areas. However, higher proportion of college productions and attractions,

which is always related to better traffic control measures and high number of walking trips, leads

to mixed effects on pedestrian crashes.

*4.2.6 Demographic characteristics*

‘Density of total population’ was the influential macroscopic variable and was positively

associated with all three target variables. The marginal effects of population density near an

intersection showed that an increase in population density will increase the frequency of motor

vehicle, bicycle and pedestrian crashes by 2.48, 0.08, and 0.12. This agrees with previous studies

as a larger population is always consistent with more opportunities in terms of crash exposure

(Lee et al., 2015; Mitra and Washington, 2012; Pulugurtha and Sambhara, 2011). In addition, the

proportion of population between age 16 and 64 near an intersection was found to have negative

effects on pedestrian crashes. The frequency of pedestrian crashes will decrease by 0.04 with a

percentage increase in proportion of population aged 16-64. This may be due to middle-aged

people walking less in comparison to young and/or old people, as well as having better ability in

avoiding crash risk (Huang et al., 2010).

*4.2.7 Commute behaviors*

A percentage increase in proportion workers commuting by public transportation near an

intersection was associated with a 0.50 decrease in motor vehicles and a 0.10 increase in bicycle

crashes, indicating that the public transportation had opposite effects on motor vehicles and

bicycle crashes. In addition, ‘proportion of workers commuting by walking’ had significant and

positive associations with pedestrian crash occurrence. The number of pedestrian crashes will

increase by 0.14, with a percentage increase in proportion workers commuting by walking. This

result is not surprising since walking is always associated with the exposure of pedestrian

crashes.

## 5. Conclusions and recommendations

This paper sought to examine the effects of omitted macroscopic factors in crash-frequency

models by transportation modes at intersections. For this purpose, several separate

intersection-level crash-frequency model for motor vehicle, bicycle and pedestrian modes were

developed. Road characteristics related to geometric design and regulatory/control attributes and

traffic characteristics of the intersection entities, as well as macroscopic factors including trip

production/attraction, demographic and socio-economic characteristics and commute behaviors

at TAZ level surrounding the intersection, were used as explanatory variables. Those data

extracted for 279 intersections located in Hillsborough County, Florida, USA, were used for

model development.

The empirical analysis revealed a number of interesting findings. First, omission of

macroscopic variables would result in biased estimation of retained microscopic variables.

Results for marginal effects of traffic volumes and road features showed significant differences

between in the base model and the full model with macroscopic variables. For example, the safety

effect of minor approach AADT on motor vehicle and bicycle crashes are biased downwards by

16.8% and 17.6% in the absent of macroscopic variables.

Second, macroscopic variables had potential effects in tracking the heterogeneity of certain

risk factors. The results in the base model using RPNB approach showed that the safety effect of

‘presence of traffic signal’ was best fit with a normally distributed random parameter suggesting

‘presence of traffic signal’ had heterogeneous effects across intersections; while this variable

became no statistically significant in models with macroscopic variables. This implied that the

heterogeneous effects of ‘presence of traffic signal’ on motor vehicles crashes could be mostly

captured by these macroscopic variables, at least for the Hillsborough dataset.

Third, model comparison using log likelihood at convergence suggested that considering

macroscopic variables was vital in elevating the model performance. In addition, the values of

PRV were further calculated to assess the explanatory power of macroscopic variables. Results

showed the values of PRV in bicycle and pedestrian crash-frequency model were much higher

than in motor vehicle models, indicating that integrating macroscopic factors played a more

important role in developing non-motorized crash-frequency model than in developing motor

vehicle models.

Fourth, comparing model outputs developed based on 0.25 mile, 0.5 mile and 1 mile buffer

width data, models with macroscopic variables of 0.25 mile buffer width had the lowest AIC and

highest PRV, conversely, the models with macroscopic variables of 1 mile buffer width had the

highest AIC and lowest PRV in all three types of crash-frequency models by modes. Thus a

relatively smaller buffer width to extract macroscopic factors around the intersection would

provide a better estimate.

Finally, macroscopic factors of the surrounding zone of an intersection, such as ‘proportion

of home-based productions and attractions’, ‘proportion of college productions and attractions’,

‘density of total population’, ‘proportion of population between age 16 and 64’, proportion of

workers commuting by public transportation and ‘proportion of workers commuting by walking’,

were demonstrated to have significant effects on intersection crashes; while these variables are

always ignored in traditionally micro-level (e.g., intersections and segments) crash frequency

model. This indicated that not only traffic volumes and road features but also macroscopic factors

should be considered in estimating crash risk and identifying crash-prone locations.

The topic of integrating macroscopic factors in intersection/segment-level crash-frequency

model is emerging. This study has great research potential in pro-actively predicting crash risk

and identifying suitable countermeasures to reduce the crashes at “new” intersections/segments

as well as intersections/segments near “new” development. Nevertheless, several limitations

should be noted for this study. First, it is worthwhile to apply this model to other intersections

and regions in order to investigate spatial transferability of the calibrated models. In addition,

there may be some correlations among crash frequency by different transportation modes within

intersections, which is caused by some unobserved influential factors. Therefore, a simultaneous

model accounting for correlations of crashes among transportation modes, such as the

multivariate model, will be further explored to investigate this issue.

## Acknowledgements

This work was jointly supported by: 1) Natural Science Foundation of China (No.71371192,

No. 71561167001); 2) the Research Fund for Fok Ying Tong Education Foundation of Hong Kong

(142005); and 3) Fundamental Research Funds for the Central Universities of CSU (No.

2016zzts050). We would like to thank Dr. Mohamed Abdel-Aty at the University of Central

Florida and the Florida Department of Transportation for providing the data.

## References

Abdel-Aty, M., Siddiqui, C., Huang, H., Wang, X., 2011. Integrating trip and roadway

characteristics to manage safety in traffic analysis zones. Transportation Research Record:

Journal of the Transportation Research Board 2213, 20-28.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies

with random-parameters count models. Accident Analysis and Prevention 41, 153-159.

Christoffel, T., Gallagher, S., 1999. Injury Prevention and Public Health: Practical Knowledge,

Skills, and Strategies. Aspen Publishers Inc., Gaithersburg, MD.

Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of

traffic analysis zones: Assessing the spatial proximity effects. Accident Analysis and

Prevention 82, 192–198.

Dong, N., Huang, H., Lee, J., Gao, M., Abdel-Aty, M., 2016. Macroscopic hotspots identification: A

Bayesian spatio-temporal interaction approach. Accident Analysis and Prevention 92, 256-264.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida:

Bayesian spatial modeling. Transportation Research Record: Journal of the Transportation

Research Board 2148, 27-37.

Huang H., Song B., Xu P., Zeng Q., Lee, J., Abdel-Aty, M., 2016. Macro and micro models for zonal

crash prediction with application in hot zones identification. Journal of Transport Geography, 54, 248-256.

Jacobsen, P.L., 2003. Safety in numbers: More walkers and bicyclists, safer walking and bicycling.

Injury Prevention 9, 205-209.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at

intersections in Florida. Accident Analysis and Prevention 37, 775-786.

Lee, J., Abdel-Aty, M., Jiang X., 2015. Multivariate crash modeling for motor vehicle and

non-motorized modes at the macroscopic level. Accident Analysis and Prevention 78, 146-154.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and

assessment of methodological alternatives. Transportation Research Part A 44, 291-305.

Mitra, S., Washington, S., 2012. On the significance of omitted variables in intersection crash

modeling. Accident Analysis and Prevention 49, 439-448.

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment,

pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections.

Accident Analysis and Prevention 43, 1624-1634.

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier

and future directions. Analytic Methods in Accident Research 1, 1-22.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical

analysis of highway accident data. Analytic Methods in Accident Research 11, 1-16.

NHTSA, National Highway Traffic Safety Administration, FARS Data, 2013.

www-fars.nhtsa.dot.gov/Main/index.aspx

Pulugurtha, S.S., Sambhara, V.R., 2011. Pedestrian crash estimation models for signalized

intersections. Accident Analysis and Prevention 43, 439-446.

Quddus, M., 2008. Modeling area-wide count outcomes with spatial correlation and

heterogeneity: an analysis of London crash data. Accident Analysis and Prevention 40, 1486-1497.

Raudenbush, S.W., Bryk, A.S., 2002. Hierarchical linear models: Applications and data analysis

methods, 2nd ed. Sage, Thousand Oaks.

Ryb, G.E., Dischinger, P.C., Kufera, J.A., Soderstrom, C.A., 2007. Social, behavioral and driving

characteristics of injured pedestrians: A comparison with other unintentional trauma patients.

Accident Analysis and Prevention 39, 313-318.

Strauss, J., Miranda-Moreno, L.F., Morency, P., 2013. Cyclist activity and injury risk analysis at

signalized intersections: a Bayesian modeling approach. Accident Analysis and Prevention 59, 9-17.

Strauss, J., Miranda-Moreno, L.F., Morency, P., 2014. Multimodal injury risk analysis of road users

at signalized and non-signalized intersections. Accident Analysis and Prevention 71, 201-209.

Ukkusuri, S., Miranda-Moreno, L.F., Ramadurai G., Isa-Tavarez, J., 2012. The role of built

environment on pedestrian crash frequency. Safety science 50, 1141-1151.

Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate

crash frequencies by severity, number of vehicles involved, collision and location type.

Accident Analysis and Prevention 59, 309-318.

Wang, J., Huang, H., 2016. Road network safety evaluation using Bayesian hierarchical joint

model. Accident Analysis and Prevention 90, 152-158.

Xie, K., Wang, X., Huang, H., Chen, X., 2013. Corridor-level signalized intersection safety analysis

in Shanghai, China using Bayesian hierarchical models. Accident Analysis and Prevention 50, 25-33.

Xu, P., Huang, H., Dong, N., Abdel-Aty, M., 2014. Sensitivity analysis in the context of regional

safety modeling: Identifying and assessing the modifiable areal unit problem. Accident

Analysis and Prevention 70, 110–120.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: Random parameter versus

geographically weighting. Accident Analysis and Prevention 75, 16-25.

Xu, P., Huang, H., Dong, N., Wong, S., 2016. Revisiting crash spatial heterogeneity: a Bayesian

spatially varying coefficients approach. Accident Analysis and Prevention, DOI: 10.1016/j.aap.2016.10.015.

Zeng, Q., Huang, H., 2014. Bayesian spatial joint modeling of traffic crashes on an urban road

network. Accident Analysis and Prevention 67, 105-112.